DOCUMENT RESUME

ED 109 260 TM 004 729

AUTHOR Klein, Stephen P.; And Others
TITLE A Progress Evaluation of Four Bilingual Children's Television Shows.
PUB DATE Apr 75
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D. C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Academic Achievement; Analysis of Covariance; *Biculturalism; *Bilingual Education; Bilingual Students; Criterion Referenced Tests; *Educational Television; Elementary Education; *Elementary School Students; Grade 1; Program Effectiveness; *Program Evaluation; Self Concept; Spanish Speaking; Test Construction

ABSTRACT

An evaluation of a bilingual education TV series was conducted involving 6-year-old English speaking, Spanish speaking, and bilingual children at four sites. Children were assigned to control and experimental groups with the latter group seeing four 30 minute shows. A pretest-posttest design was employed with the pretest serving as the covariate in the analyses of the data. Results indicated statistically but not educationally significant effects of the shows for certain objectives. The implications of these findings and the factors that may have influenced them are discussed. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED109260

A PROGRESS EVALUATION OF FOUR BILINGUAL CHILDREN'S TELEVISION SHOWS

Stephen P. Klein
Ralph Hoepfner
Jacqueline P. Kosecoff

Center for the Study of Evaluation
UCLA Graduate School of Education

TM 004 229

Paper presented at the Annual Meeting of the National Council on Measurement in Education, Washington, D.C., March 31 through April 2, 1975.

A PROGRESS EVALUATION OF FOUR BCTV SHOWS

Bilingual Children's Television (BCTV) is a federally and privately funded project that is in the process of developing a daily television series. This series is being constructed as a supplement to the normal instructional services provided by schools and is aimed at early elementary school children who speak English and/or Spanish. In brief, there are five general goals which underlie BCTV's objectives. These are:

1. To promote cultural pluralism through an understanding and appreciation of Latin-American culture, heritage, values and mores.

2. To provide the Latin-American child with an experience in which his or her language and culture predominate, for the purpose of enhancing his or her self-concept, encouraging his or her continued or renewed interest in his or her background, and help to bridge the gap that may exist between home and school.

3. To provide non-Spanish speakers with an opportunity to become familiar with the Spanish language and to help all viewers recognize the advantage of speaking more than one language.

4. To aid the viewer in developing the communication and problem solving skills necessary to function successfully in his or her environment.

5. To present selected information designed to lead to the development of concepts in the areas of Human Relations, Food and Nutrition, Natural Environment, Energy, and Man-made Objects.

Each show in the series is composed of a group of program segments that are focused on a central theme, such as the "need for practice," which in turn stem from one of the five areas listed in objective number five. A monolingual child (or adult) can readily understand what is happening in a given program segment by watching the action and/or listening to what is

said. Each show lasts for 30 minutes and involves animation, film excerpts, and "live action" in its several segments.

In an evaluation of the two prototype (pilot) shows for the BCTV series, it was ascertained that the shows' basic format was generally entertaining and attractive; and, the content was understood by the children in the target audience (BCTV, 1973). The evaluation of the prototype shows did not examine whether they were achieving their instructional objectives since it was anticipated that significant learning effects and attitude changes would not become evident until a child had had an opportunity to view a substantial portion of the whole series. However, BCTV desired some preliminary information with respect to the attainment of objectives as soon as several shows were essentially ready for distribution. This need arose from the desire to have a sound basis for making decisions regarding possible modifications in the general developmental strategy. Thus, the specific purpose for the present evaluative-research was to assess the extent to which four of the initially developed shows were making progress in achieving BCTV's objectives.

Procedures

The basic design for this study involved the use of individually administered criterion referenced pretests and posttests to semi-intact groups of control and experimental children in the target populations. The experimental group saw each of the four shows twice while the control group did not see any of the shows.

Subjects

Six year old children in the first grade who spoke English and/or Spanish at each of four sites were used in this study. The four sites (Albuquerque, New York City, Salt Lake City, and San Jose) were selected to provide

Table 1

CELL FREQUENCIES

		ALBUQUERQUE		NEW YORK		SALT LAKE CITY		SAN JOSE		TOTALS	
		Con.	Exp.	Con.	Exp.	Con.	Exp.	Con.	Exp.		
FEMALE	ENG.	8	10	7	10	31	40	21	18	145	
	SPAN.	3	2	7	6	2	2	5	2	29	248
	BIL.	14	14	8	14	0	3	7	4	64	
MALE	ENG.	11	9	13	16	32	19	21	19	140	
	SPAN.	0	4	6	9	1	1	2	1	24	228
	BIL.	25	16	8	16	0	1	7	1	74	
		61	55	49	71	66	66	63	45		
		116		120		132		108			

N=476

geographic and cultural variety. The children at each site were assigned to experimental and control conditions in randomized clusters in order to minimize assignment bias and yet not interfere unduly with the schools' routines.

At the beginning of the pretest, the language dominance of each child was determined (i.e., English, Spanish, or Bilingual). The standardized procedures used for this purpose are presented in Appendix A. The evaluation plan called for 10 students in each of the design's 48 cells (i.e., 2 groups X 4 sites X 3 categories of language dominance X 2 sexes). The actual number of subjects in each cell who completed both the pretest and posttest are presented in Table 1. The n's for children in the experimental group in this Table refer to just those children who saw all four shows twice.

Evaluation Test Construction

In order for scores on the evaluation tests to be veridical reflections of the effects of the shows, each item of the test was criterion-referenced to a particular goal of the shows. Representatives of Bilingual Children's Television, Inc., Educational Evaluation Associates, and the Institute of Cultural Pluralism viewed the shows and then consensually constructed test items that could reasonably be assumed to be sensitive to the effects of the shows. Specific concerns underlying the items so developed were:

1. Age Appropriateness. The items should be appropriate, both in content and difficulty, for children at age six. This meant that the conservation items had to be sufficiently difficult so that not all children would get the items correct, regardless of the shows. In addition, the language, both English and Spanish, had to be carefully selected to be well within the receptive repertoire of six-year old children.

2. Instructional Independence. Although long term conceptual development is the ultimate goal of most educational programs, including the BCTV series, the time constraints on the evaluation called for the assessment of immediate learning of the shows' content and intent. The generalizability of the learning was to be assured by the construction of items with the same content as the shows, but with different formats and applications. In this way, scores on the items would not represent merely a short-term rote memory effect. Thus, item presentations were constructed so as to be significantly different from presentations used in the actual shows. By this means, it could be assurred that the child who correctly answered a test item was in fact transferring what he/she had learned from the show to a novel setting which differred in method of presentation while still maintaining the essential educational intent of the show.

3. Bias. To avoid cultural and sex biases in the items, names of the characters were selected to be common to both English and Spanish-speaking cultures; and, the characters were balanced as to sex membership. The test development team also represented diversity in language dominance, sex, and cultural backgrounds.

4. Objectives Measured. The funding and time constraints on this study did not allow for a complete assessment of all of the BCTV goals. Thus, it was decided to sample a range of objectives from the most general to the most specific in three general areas of concern, namely: Labelling, Concepts, and Attitudes. Within these areas, the specific objectives chosen were the ones that were emphasized the most across all four shows used in this study.

5. Practicality. There were a number of unique constraints that had to be taken into consideration in the design of administration of the tests,

such as the number of shows available for distribution at the inception of the study, the availability of classroom time for viewing the shows towards the end of the school year, the availability of a cadre of bilingual examiners at a given site as well as classrooms containing adequate numbers of Spanish speaking or bilingual children (so that normal school operations would not be disrupted), and related concerns.

The Evaluation Test

The evaluation test was constructed in parallel English and Spanish forms. This was done so that the child took the test in the language in which he/she was dominant. In the case of a bilingual child, that child was given the test form corresponding to the language in which he/she had expressed a preference to be tested.

The format for most of the items involved the examiner's reading of the item in the test booklet while the child followed illustrations in a separate booklet common to both the English and Spanish forms. The child then indicated his/her response verbally or by pointing to a picture or object in the separate booklet. The attitude items were an exception to this procedure and are discussed in more detail below.

In order to stay within the desired 10 to 20 minute administration time period, the test was limited to 12 items. Some of these items, however, yielded several points of credit. Copies of the complete set of test and scoring materials for the English and Spanish Forms are contained in Appendices B and C, respectively. Appendix A contains the booklet common to both forms.

The three major areas for which test items were constructed and the specific objectives within each area assessed were as follows: "Labelling" (Number Identification), "Concepts" (Number Conservation and Content Theme),

and "Attitudes" (Towards English and Spanish languages). A description of the five subtests used to measure these objectives is presented below:

1. Number Identification (Score range: 0 to 8 points). Periodically within each show, a segment was concerned with learning the name of a numeral was also embedded into aspects of short story sequences. The three items used to measure this area (items C, E, and J) involved naming the numeral illustrated in both languages and identifying the numeral spoken in the non-dominant language. The numerals used for these items were those that were emphasized in the shows.

2. Number Conservation (Score range: 0 to 2 points). The numeral emphasized within a given show also was visually permuted in many ways to illustrate, in an indirect manner, that it conserved its number quality throughout the permutations. The two items used to measure this area (A and G) involved the child recognizing that the number of objects in a situation remained the same even though they had been moved several times. The numerals used for these items were among those that were emphasized in the four shows employed in this study.

3. Thematic Content (Score range: 0 to 20 points). The basic underlying theme of each show focused upon a specific topic. The four topics represented by the shows used in this study were concerned with energy and work, pollution and ecology, the need for practice, and the varied uses of cloth. The five items used to measure this area (B, D, F, H, and I) involved a variety of tasks, such as identifying skills that did or did not need practices to develop and identifying pollutants in a nature scene.

4. Positive Bicultural Attitudes (Score range: 0 to 3 points per language). The procedure employed to assess attitude change was an adaptation of the method used by Lambert (Lambert, 1967; Lambert, Frankel, & Tucker,

1966). This technique is a variation of the semantic differential approach to attitude assessment. It involves the child hearing a short discourse in one language and then rating the speaker on a set of personal characteristics. The child then hears the same speaker with the same discourse, but in the other language. Since speaker and content are the same, differences in the two sets of ratings of personal characteristics are presumed to be attributable to attitudes held by the child toward the languages and their concommitant cultures.

In the present study, an uneventful and common passage was selected for the discourse. Two audio tapes were constructed. The first tape contained a female speaker saying the passage in Spanish and then in English. The second tape contained a male speaker saying the passage in English and then in Spanish. Female examinees heard the female speaker and male examinees heard the male speaker. Upon hearing the speaker, the child pointed to either of two words on each of the following four dimensions: short-tall (practice item, not scored); good-bad, dumb-smart, good looking-ugly.

Test and Show Administration

With few exceptions, the evaluation schedule was pretest on a Friday, presentation of the shows to the experimental group on the following Monday through Thursday, and then posttest on Friday. Each show was shown twice to the experimental group in a counterbalanced manner so that a given show was seen both in the morning and the afternoon and there was at least a one full day separation between each viewing of a given show. The shows were presented by means of closed circuit or portable video equipment.

The testing was done on an individual basis with examiners who were fluent in both English and Spanish. Each examiner used a portable cassette recorder to play the evaluation test tapes of the male and female speakers to

male and female children, respectively. In general, the examiners did not know whether a given child was in the control or experimental group; and, each examiner tested children in both groups.

A fully bilingual coordinator was assigned to each site. The coordinators' responsibilities included establishing liaisons between the schools and the independent evaluation agency, presenting the shows, training and supervising the examiners, keeping a log of student reactions to the shows and any anomalies in the testing procedures, and related functions. A summary of the coordinators' logs are presented in Appendix E.

Analysis Procedures

The data analysis plan for this study called for an analysis of covariance (ANOCOV) for each of the five dependent variables; i.e., the subtest scores for number identification, number conservation, thematic content, positive attitude toward English Language, and positive attitude toward Spanish Language. Each of the ANOCOV's was to have involved the following four factors: Treatment Group (control vs. experimental), Language Dominance (English, Spanish, and Bilingual), Sex, and Geographic Site (Albuquerque, New York City, Salt Lake City, and San Jose). The corresponding pretest score for each posttest dependent variable was to be used as the covariate to adjust for any possible differences between children in their ability to answer the evaluation test questions.

As may be seen from an inspection of Table 1, there were marked differences in the obtained cell sizes even though the total number of children on which there was complete data (N=476) was essentially the same as planned. The discrepancies between planned and actual sizes were largely due to differences among the number of children in each of the three categories of language dominance. This marked disproportionality came about as a result of

numerous systematic and extraneous factors, such as having to drop 25 bilingual children at Salt Lake since they took the pretest in Spanish and the posttest in English. Similarly, about 30 students were lost at the San Jose site due to an epidemic of mumps! Thus, it was no longer feasible or appropriate to analyze the complete factorial design for each dependent variable. In the place of the planned set of analyses of the complete design, a series of ANOCOV's with unequal cell sizes were employed to test the four main effects and the two factor interactions involving the treatment group factor. All other interactions were pooled into the error term. This approach allowed maximum analytical precision with respect to the primary issues, i.e., did the experimental group do better than the controls; and if they did, was this trend consistent across sites, sexes, and categories of language dominance?

Results

A summary of the results of the five analyses of covariance are presented in Table 2. These results indicated a significant difference between the treatment groups in two of the three areas of concern, namely: Labelling and Concepts. An inspection of the adjusted posttest means scores presented in Table 3 indicates that the experimental group performed better than the control group in both instances. Thus, the TV shows did have the desired effect in these two areas.

An examination of Table 2 further reveals that there were no significant interactions between the treatment group factor and the other three factors. In other words, the effect of the treatment was consistent across sites, sexes, and language dominance groups.

The results presented in Table 2 also indicate that there were systematic differences between sites and between language groups in how much they improved

Table 2

Summary of Results of the Analysis of Covariance[1]

Source of Variation	df	Labelling: Number Identification MS	Labelling: Number Identification F	Concepts: Number Conservation MS	Concepts: Number Conservation F	Concepts: Content Theme MS	Concepts: Content Theme F	Attitudes: Towards English MS	Attitudes: Towards English F	Attitudes: Towards Spanish MS	Attitudes: Towards Spanish F
Site	3	1.10	.56	2.50	4.38**	96.18	14.10**	1.97	4.66**	.21	.32
Group	1	12.62	6.46**	.12	.22	173.79	25.48**	.01	.03	1.43	2.20
Sex	1	11.15	5.70*	.20	.36	2.09	.31	.09	.21	7.21	11.06**
Language	2	3.16	1.62	2.43	4.26*	37.57	5.51**	.26	.62	1.57	2.41
Group X Site	3	.27	.14	.37	.65	8.16	1.20	.76	1.80	1.44	2.21
Group X Sex	1	.17	.09	.04	.08	.37	.05	.44	1.03	1.18	1.81
Group X Lang.	2	1.75	.90	.18	.31	16.61	2.44	.32	.75	.65	1.00
Error	461	1.96		.57		6.82		.42		.65	

* = .05; ** = .01 Level of Signficance

1. The covariate for each posttest score was its corresponding pretest score.

Table 3

Means and Standard Deviations for Each Group[1,2]

Group		Labelling	Concepts		Attitudes	
		Number Identification (0 to 8 pts.)	Number Conservation (0 to 2 pts.)	Content Theme (0 to 20 pts.)	Towards English (0 to 3 pts.)	Towards Spanish (0 to 3 pts.)
Control	Mean	6.62	1.23	10.19	2.63	2.44
	SD	2.04	.79	3.25	.76	.97
Experimental	Mean	6.97	1.21	11.06	2.59	2.37
	SD	1.89	.80	3.12	.78	1.02

1. The table contains the adjusted posttest mean.

2. The table contains the average of the pretest and posttest standard deviations.

in performance from pretesting to posttesting (i.e., irrespective of whether or not they were in the experimental or control group). An inspection of the cell means revealed that the differences between language groups was largely due to the English dominant children gaining more than the others. The differences between sites also appeared to be confounded with this trend in that the sites that showed the largest gains were also the ones with proportionately more English dominant children.

In two instances, there was a significant difference between males and females, namely, girls improved more than boys in number identification and in developing positive attitudes towards Spanish. However, none of these gains was in any way related to the treatment; and, therefore, of little interest to the major purpose of this study.

Discussion

The foregoing results indicated that the presentation of the shows did have the desired impact in the areas of labelling and concepts, but not in the area of the development of positive attitudes. The assessment of labelling skills was limited to number identification, i.e., could a child learn the word for a number in another language? Although a given number was emphasized in each of the four shows, it was not the same number. Thus, even though a child in the experimental group saw each show twice, the degree of exposure to each number was still relatively brief (particularly with respect to the memory span of a six year old child). Despite this handicap, the experimental group did perform better than the controls; although an inspection of Table 3 indicates that the absolute size of this advantage is not especially impressive.

Within the area of Concepts, the experimental group did better than the controls on one of the two dependent variables, namely, content theme. As

in the case of number identification, the absolute size of this advantage (about one point out of 20) was statistically significant but not particularly large from an educational point of view. There are several reasons which may account for this kind of finding. For example, as noted in the introduction, learning effects may only become truly evident after the children have seen a substantial portion of the total set of 65 shows. The reason for this is that such exposure would provide adequate reinforcement since there is similar content across the full set of shows. Further, the shows are not didactic. They endeavor to get the labels, ideas, and attitudes communicated in an indirect and a child-appealing manner so that the shows are not perceived as just another extension of "school." An additional limiting factor is that it is generally difficult to get reliable test data from six year old children even when they are tested in the language in which they are dominant. The fact that 25 students at one site "switched" their language dominance in one week is illustrative of this problem. This kind of instability (unreliability) tends to mitigate against finding significant differences since it diminishes the sensitivity of the test questions to uncover real differences between children. Thus, given all these limiting factors, the finding of even small consistent differences between the experimental and control groups is indeed noteworthy.

The failure to find significant differences across treatment groups in the measure of number conservation also may be attributable to some of the limiting variables noted above as well as to certain unique factors. One of these factors was that there were relatively few points assigned to measuring this concept and this in turn limited the degree to which differences might be found (i.e., via the limitation of the available systematic variance). Siegel (1969), Piaget, and others who have experienced difficulty in

educating children with respect to "conservation" also have come to the conclusion that this concept may not be directly trainible; thus, it would almost be presumptive to expect that the indirect methods employed in the BCTV shows to achieve this end would be successful. If this is indeed the case, then BCTV might well reconsider continued inclusion of number conservation segments in its shows. On the other hand, the segments within a show dealing with number conservation also focus on the same number that is emphasized by that show for the purposes of number identification; and, it may be argued that the number conservation segments provide valuable reinforcement for the learning of the labels. Subsequent research on this issue is needed, especially given the amount of show time devoted to number conservation segments.

The lack of significant differences in the attitude area may be attributable to many of the same factors as noted above, such as score unreliability and limited variance, as well as to certain special problems associated with attitude measurement. For example, most of the children "topped out" on both the English and Spanish scales in the sense that they were quite positive about both "speakers." This finding may be more an artifact of the assessment technique with young children than any true indication of their attitudes towards persons who speak these languages; or, it may be that such children have not even developed attitudes that discriminate between English and Spanish. This is an important consideration for BCTV since it is one of the primary goals of BCTV to develop such attitudes. It was also recognized at the inception of this study that finding differences in this area after only two showings of four shows was quite unlikely; however, given the central importance of this area, it was deemed necessary to make an attempt to find possible impacts. This point of view also led to the requirement for each site coordinator to keep a log of what occurred during both the shows

and the test sessions. A summary of these logs is presented in Appendix D and it indicates that in general the children did enjoy watching the shows and often actively engaged in activities associated with them, such as singing the songs.

Conclusions

The results of this study indicated that the four shows were effective in improving student performance in the areas of number identification and content theme. No significant differences were found between the two groups in the areas of number conservation and the development of positive attitudes towards English and Spanish. The failure to find a positive effect on the attitude measures appeared to be due at least partly to the curtailed range of scores on these measures. Thus, no conclusions may be drawn about the possible effectiveness of the shows with respect to their impact on student attitudes. An examination of the coordinator's logs does reveal, however, that the students usually enjoyed the shows and it is presumed that such involvement may eventually have carry-over effects in the sense of improving student attitudes. Finally, it should be noted that the present study only involved the presentation of four different shows (even though each one was shows twice). This point is important since the philosophy underlying the BCTV series is that the major beneficial effects of the shows will only become fully evident after the child has seen a substantial portion of them.

In conclusion, it appears that the shows are generally making satisfactory progress in achieving at least some of their general goals.